ELSEVIER

Operations Research Letters 20 (1997) 109–118

operations research letters

Approximation-assisted point estimation[1]

Barry L. Nelson[a], Bruce W. Schmeiser[b,*], Michael R. Taaffe[c], Jin Wang[d]

[a] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, USA
[b] School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA
[c] Department of Operations and Management Science, Faculty of Scientific Computation, University of Minnesota, Minneapolis, MN 55455, USA
[d] Department of Mathematics and Computer Science, Valdosta State University, Valdosta, GA 31698-0040, USA

Received 1 May 1994; revised 7 July 1995

Abstract

We investigate three alternatives for combining a deterministic approximation with a stochastic simulation estimator: (1) binary choice, (2) linear combination, and (3) Bayesian analysis. Making a binary choice, based on compatibility of the simulation estimator with the approximation, provides at best a 20% improvement in simulation efficiency. More effective is taking a linear combination of the approximation and the simulation estimator using weights estimated from the simulation data, which provides at best a 50% improvement in simulation efficiency. The Bayesian analysis yields a linear combination with weights that are a function of the simulation data and the prior distribution on the approximation error; the efficiency depends upon the quality of the prior distribution. © 1997 Published by Elsevier Science B.V.

Keywords: Simulation; Monte Carlo; Biased estimation; Control variates

1. Introduction

We consider the analysis of stochastic systems, with the purpose of evaluating a performance measure θ. In a number of real problems, the practitioner has available both an approximation of θ that provides a constant $\tilde{\theta}$ and a stochastic simulation experiment that provides a point estimator $\hat{\theta}$. Our problem is to combine $\tilde{\theta}$ and $\hat{\theta}$ to obtain a single point estimator.

We make two context assumptions. First, the quality of $\tilde{\theta}$ is unknown; the analyst, however, respects its ability to approximate θ. Second, the point estimator is unbiased and its variance, $\sigma^2_{\hat{\theta}} \equiv \mathrm{var}(\hat{\theta})$, is estimated by an observable $\hat{\sigma}^2_{\hat{\theta}}$.

For example, consider expected waiting time at a particular queueing-network node. The constant $\tilde{\theta}$ could be a numerical approximation (e.g., Whitt, 1983) or an expert's opinion. The point estimator $\hat{\theta}$ could be the sample average of the customer waiting-time data from a simulation experiment with standard error $\sigma_{\hat{\theta}}$. Our problem is then to combine the approximation or opinion $\tilde{\theta}$ and the simulation result $\hat{\theta}$ using the estimated standard error $\hat{\sigma}_{\hat{\theta}}$.

Faced with such a context, the analyst sometimes chooses between $\tilde{\theta}$ and $\hat{\theta}$. The idea is to use $\tilde{\theta}$ if it

* Corresponding author. Tel.: +1 317 494 5422; fax: +1 317 494 1299; e-mail: schmeise@ecn.purdue.edu.

[1] This work was supported by NSF Grant DDM-93-00058 and the Minnesota Supercomputer Institute. The authors thank David Goldsman for insightful comments and detailed suggestions for improved presentation.

PII S0167-6377(96)00053-3

is "verified" by the simulation; that is, if $|\tilde{\theta}-\hat{\theta}|$ is *small*, choose to report $\tilde{\theta}$, otherwise choose $\hat{\theta}$. This *binary choice* is essentially classical Neyman–Pearson hypothesis testing, with the null hypothesis $H_0: \theta=\tilde{\theta}$ and the alternative $H_1: \theta \neq \tilde{\theta}$. We investigate binary choice, including methods for estimating the procedure parameter that defines *small*, in Section 2.

An alternative to binary choice is to linearly combine $\tilde{\theta}$ and $\hat{\theta}$. We investigate linear combinations, including weight estimation by minimizing mean squared error (mse), in Section 3, and by Bayesian analysis, in Section 4.

For both binary choice and linear combinations we discuss performance as measured by mse, the sum of the squared bias and the variance. The approximation $\tilde{\theta}$ is deterministic, which can be interpreted as zero variance; its error, $\tilde{\theta}-\theta$, can be interpreted as bias. The point estimator $\hat{\theta}$ is unbiased, but has variance $\text{var}(\hat{\theta})$, which is inversely proportional to the sample size n. (These properties of $\hat{\theta}$ apply widely, especially for large sample sizes.)

We show that binary choice is at best modestly effective, that it is dominated (as a practical matter) by the natural-estimator linear combination, and that the Bayesian analysis is effective or not depending upon the quality of the prior information.

2. Binary choice

Assume that only the observed values $\tilde{\theta}, \hat{\theta}$, and $\hat{\sigma}_{\hat{\theta}}$ are available and we wish to report the point estimator defined by choosing either $\tilde{\theta}$ or $\hat{\theta}$. The binary-choice family of estimators is

$$\hat{\theta}^{\text{b}}(\beta)=\begin{cases}\tilde{\theta} & \text{if } |\tilde{\theta}-\hat{\theta}|<\beta\hat{\sigma}_{\hat{\theta}},\\ \hat{\theta} & \text{otherwise.}\end{cases} \tag{2.1}$$

The single parameter β is nonnegative, either a prespecified constant or a function of the observed values $\tilde{\theta}, \hat{\theta}$ and $\hat{\sigma}_{\hat{\theta}}$. If $\beta=0$, then the simulation estimator is returned always; if $\beta=\infty$, then the numerical approximation is returned always. Based on its relationship to Neyman–Pearson hypothesis testing, a prespecified value of β in the range from one to three seems reasonable to many people. Under the alternative hypothesis $H_1: \theta \neq \tilde{\theta}$, the probability that the simulation estimator $\hat{\theta}$ is chosen is the probability of a type-II error for a given value of $\tilde{\theta}-\theta$. If $\hat{\theta}$ is the sample average of independent and normally distributed data, then a standard derivation for the two-sided hypothesis test yields

$$\text{P}(\hat{\theta}^{\text{b}}(\beta)=\hat{\theta})=1+F_T(-\beta)-F_T(\beta), \tag{2.2}$$

where $T=(\hat{\theta}-\tilde{\theta})/\hat{\sigma}_{\hat{\theta}}$ and F_T is the noncentral-t distribution function with $n-1$ degrees of freedom and noncentrality parameter $\sqrt{n}(\theta-\tilde{\theta})/\sigma_{\hat{\theta}}$.

Consider mse performance of binary-choice estimators for independent and normally distributed data. Fig. 1 shows the mse of $\hat{\theta}^{\text{b}}(\beta)$ plotted as a function of the squared approximation error $(\tilde{\theta}-\theta)^2$. Both axes are labeled in units of $\text{var}(\hat{\theta})$. The two extreme lines, which do not depend on distributional assumptions, correspond to deterministically using the simulation or the approximation. The horizontal line at a height of 1 corresponds to the simulation point estimator $\hat{\theta}=\hat{\theta}^{\text{b}}(0)$, with $\text{mse}(\hat{\theta},\theta)=\text{var}(\hat{\theta})$, since bias is zero. The 45° line, corresponding to the squared bias of the deterministic approximation $\tilde{\theta}=\hat{\theta}^{\text{b}}(\infty)$, is $\text{mse}(\tilde{\theta},\theta)=(\tilde{\theta}-\theta)^2$.

Compare $\tilde{\theta}$ and $\hat{\theta}$. The mse of the approximation $\tilde{\theta}$ grows linearly; the mse of the simulation estimator $\hat{\theta}$ is $\text{var}(\hat{\theta})$, which decreases with sample size and is not a function of $\tilde{\theta}$. The mse of $\tilde{\theta}$ is smaller than the mse of $\hat{\theta}$ if $|\tilde{\theta}-\theta|<\sigma_{\hat{\theta}}$; i.e., if the true value and approximation differ by less than one standard error. Because $\tilde{\theta}$ and θ are constants, and because $\sigma_{\hat{\theta}}$ decreases as the simulation run length increases, eventually $\text{mse}(\hat{\theta},\theta)$ decreases to a value less than $\text{mse}(\tilde{\theta},\theta)$.

The four curves in Fig. 1 correspond to the parameter values $\beta=1$, 1.6, 2, and 3. These curves indicate the continuum of mse performance as β is varied between the two extremes. Large β values perform well when the approximation $\tilde{\theta}$ is good, but substantially increase the mse when the approximation $\tilde{\theta}$ is poor. All curves corresponding to finite values of β are asymptotic to the simulation line.

Fig. 1 was drawn with simulation as the benchmark method, with its mse held constant to one. Containing the same information, and equally fundamental, the analogous Fig. 2 is scaled by $\text{mse}(\tilde{\theta},\theta)$ rather than by $\text{mse}(\hat{\theta},\theta)$. The effect is that the numerical approximation is shown as the horizontal line and simulation as the 45° line. Despite being redundant,

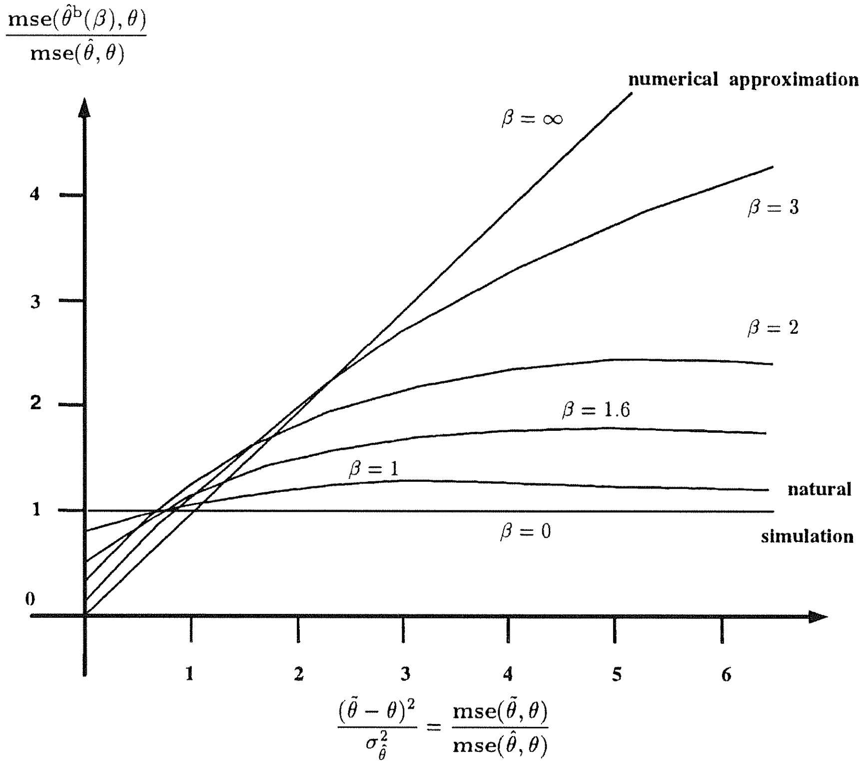

Fig. 1. For various parameter values β, the binary-choice estimator mse as a function of the approximation mse, both scaled by the simulation mse.

Fig. 2 provides additional insight by drawing the eye to situations where the approximation is better than simulation, that is, to the right side of the figure, as defined by $\mathrm{mse}(\tilde{\theta},\theta)/\mathrm{mse}(\hat{\theta},\theta) < 1$. For each value of β, the mse curve increases with simulation mse. Each curve approaches (from above) its tangent line, which passes through the origin and has slope equal to the mse of the zero-error approximation from Fig. 1. For example, the curve corresponding to $\beta = 1$ has an asymptotic slope of about 0.8.

The mse optimal value of β as a function of the bias $\tilde{\theta}-\theta$ is

$$\beta^* = \begin{cases} \infty & \text{if } |\tilde{\theta}-\theta| < \sigma_{\hat{\theta}}, \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, the optimal binary-choice estimator is the numerical approximation $\tilde{\theta}$ if and only if $\mathrm{mse}(\tilde{\theta},\theta) < \mathrm{mse}(\hat{\theta},\theta)$.

Because θ and $\sigma_{\hat{\theta}}$ are unknown, the optimal value of β must be estimated. The natural estimator, obtained by substituting the observed value $\hat{\theta}$ for θ and the sample standard error $\hat{\sigma}_{\hat{\theta}}$ for $\sigma_{\hat{\theta}}$, is

$$\hat{\beta}^* = \begin{cases} \infty & \text{if } |\tilde{\theta}-\hat{\theta}| < \hat{\sigma}_{\hat{\theta}}, \\ 0 & \text{otherwise.} \end{cases}$$

Substituting $\hat{\beta}^*$ into Eq. (2.1) for β and simplifying yields the *natural binary-choice estimator* $\hat{\theta}^{\mathrm{b}}(1)$. Therefore, although the optimal weight is a function of the approximation bias, and the estimated optimal weight is a function of the random point estimator and the random standard-error estimator, simplification yields the deterministic curve corresponding to $\beta = 1$.

The (perhaps surprising) result is that the natural binary-choice estimator is not much better than

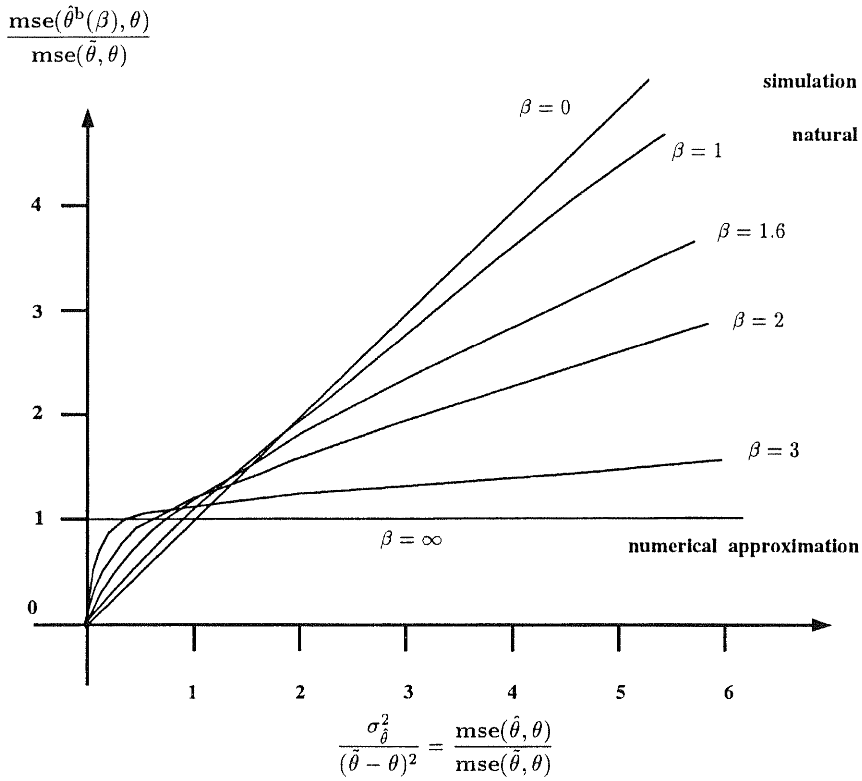

Fig. 2. For various parameter values β, the binary-choice estimator mse as a function of the simulation mse, both scaled by the approximation mse.

simulation alone. In Fig. 1, the $\beta = 1$ curve is minimized when the approximation is exact, at which point the mse is reduced by approximately 20%. The $\beta = 1$ curve is maximized when the squared approximation error is about $(\tilde{\theta} - \theta)^2 = 3\,\text{var}(\hat{\theta})$, at which point the mse is increased by about 26%. Therefore, the natural binary-choice estimator never is much better or much worse than always choosing the simulation estimator $\hat{\theta}$.

The explanation for the modest improvement is that when the approximation is good, the binary-choice rule chooses the simulation estimator only when it is unusually low or high; thus the rule filters out the good simulation point estimates. This effect can be seen in Fig. 3. The probability of choosing the simulation estimator, from Eq. (2.2), is shown in Fig. 3(a); the mse of the simulation estimator, conditional on it being chosen, is shown in Fig. 3(b).

As is consistent with intuition and Fig. 3(a), the probability of choosing the simulation estimator increases with $1/\beta$ and $(\tilde{\theta} - \theta)^2$. The natural estimator, $\beta = 1$, has the appealing property that the probability of choosing the simulation is greater than the probability of choosing the approximation if and only if the simulation mse is less than the approximation mse. In particular, $\text{P}(\hat{\theta}^b(1) = \hat{\theta}) = 0.5$ at $(\tilde{\theta} - \theta)^2 = \text{var}(\hat{\theta})$.

Fig. 3(b) shows the effect of using binary choice on the simulation mse. When the approximation is relatively accurate, the simulation has a high mse when it is chosen. Conversely, when the approximation is relatively inaccurate, the simulation has a low mse when it is chosen. This conditioning effect on the simulation mse explains why the binary-choice estimator is never much better or worse than simulation alone. In the limit as the approximation error grows large, all

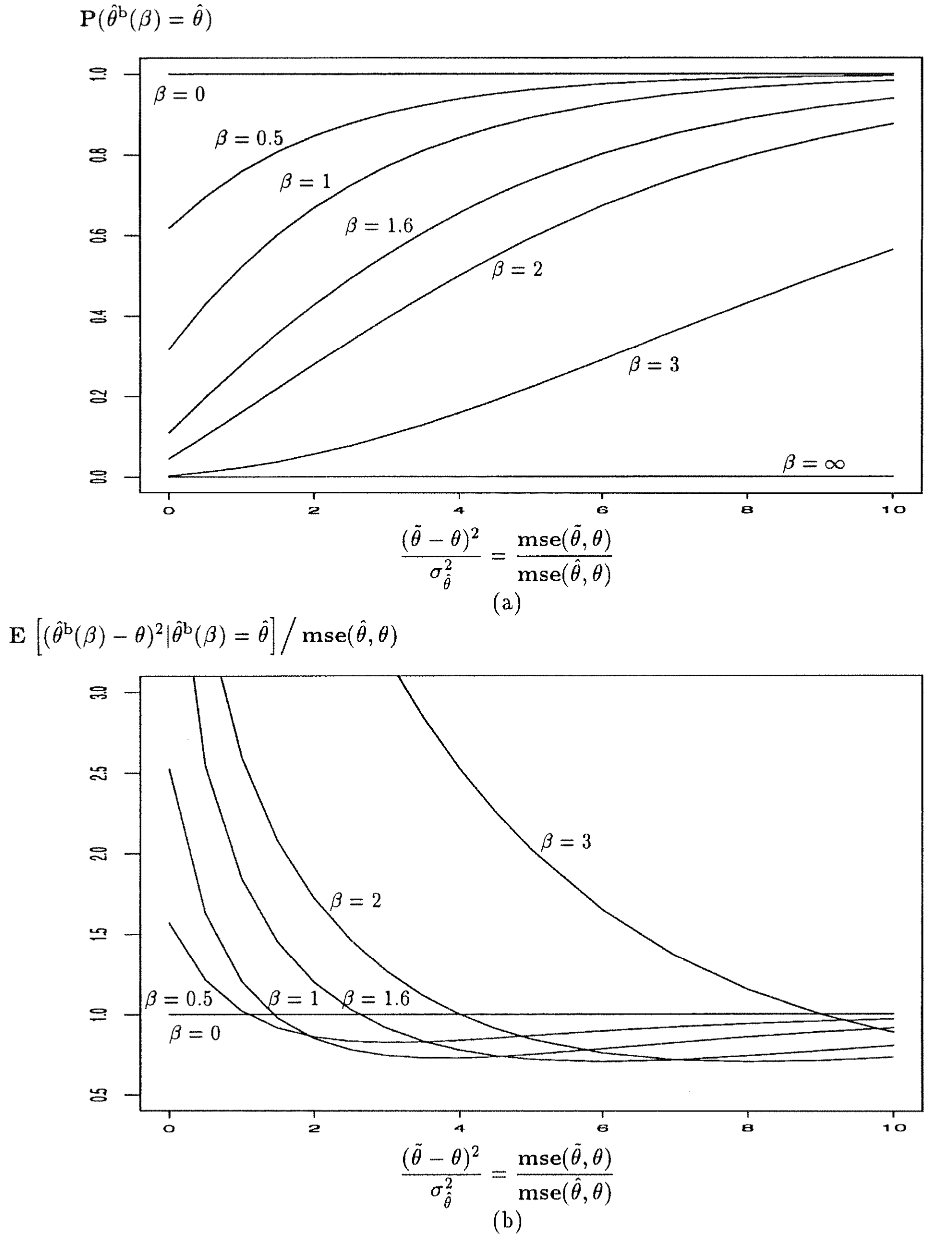

Fig. 3. Simulation performance under binary choice. (a) Probability of choosing the simulation estimator and (b) conditional simulation mse, scaled by the unconditional simulation mse.

of the simulation mse curves converge to the unconditional line from below.

In practice, the performance of binary choice is a bit worse than discussed. The derived results assume that the sample variance equals the population variance (that is, infinite degrees of freedom), and Monte Carlo results (obtained to negligible sampling error) assume samples of size thirty (that is, 29 degrees of freedom). Fig. 4, which is similar to Fig. 1, shows performance for the natural ($\beta = 1$) binary-choice

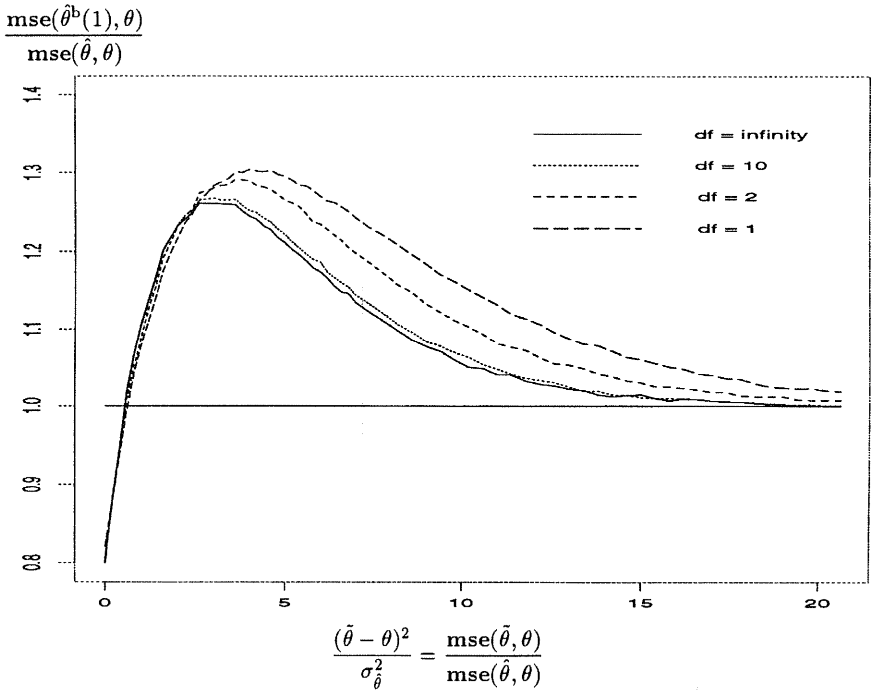

Fig. 4. For various degrees of freedom, the natural binary-choice estimator mse as a function of the approximation mse, both scaled by the simulation mse.

estimator for various degrees of freedom. The (Monte Carlo) results show surprising insensitivity to degrees of freedom when the approximation is good, and performance approaches that of simulation when the approximation is asymptotically poor. Practically significant performance degradation occurs only when degrees of freedom are quite small and the approximation error is moderate. In the worst case, one degree of freedom yields a 30% mse increase (compared to the simulation estimator $\hat{\theta}$) when the approximation mse is four or five times the simulation mse; the corresponding worst case for many degrees of freedom is about 25%.

One could also look at the sensitivity to nonnormality of the data. But the effect of data nonnormality is only through the distribution of the sample variance. If these effects are of concern, they can be controlled by grouping the observations into ten or more batches and averaging within each batch. The batch averages are asymptotically normal and the effect of fewer degrees of freedom is minor, as shown in Fig. 4.

3. MSE-optimal linear combination

Again, assume that only the observed values $\tilde{\theta}$, $\hat{\theta}$, and $\hat{\sigma}_{\hat{\theta}}$ are available. Now the analyst wishes to report the point estimator defined by the linear combination

$$\hat{\theta}(\alpha) = \alpha\hat{\theta} + (1-\alpha)\tilde{\theta},$$

which can be viewed as adjusting the approximation with the simulation, $\hat{\theta}(\alpha) = \tilde{\theta} - \alpha(\tilde{\theta} - \hat{\theta})$, or as adjusting the simulation with the approximation, $\hat{\theta}(\alpha) = \hat{\theta} + (1-\alpha)(\tilde{\theta} - \hat{\theta})$.

Writing mse as the squared bias plus the variance and as a function of the weight α yields

$$\text{mse}(\hat{\theta}(\alpha),\theta) = (1-\alpha)^2(\tilde{\theta}-\theta)^2 + \alpha^2\,\text{var}(\hat{\theta}) \qquad (3.1)$$

or equivalently

$$\text{mse}(\hat{\theta}(\alpha),\theta) = (1-\alpha)^2\text{mse}(\tilde{\theta},\theta) + \alpha^2\,\text{mse}(\hat{\theta},\theta).$$

The linear-combination mse is less than the simulation mse whenever

$$\alpha \in \left(\frac{\text{mse}(\tilde{\theta},\theta) - \text{mse}(\hat{\theta},\theta)}{\text{mse}(\tilde{\theta},\theta) + \text{mse}(\hat{\theta},\theta)}, 1 \right).$$

Minimizing the mse in Eq. (3.1) with respect to α yields the mse-optimal linear-combination weight

$$\alpha^* = \frac{(\tilde{\theta}-\theta)^2}{(\tilde{\theta}-\theta)^2 + \text{var}(\hat{\theta})} = \frac{\text{mse}(\tilde{\theta},\theta)}{\text{mse}(\tilde{\theta},\theta) + \text{mse}(\hat{\theta},\theta)}.$$

The unusual interpretation of $\hat{\theta}$ as a biased estimator of $\tilde{\theta}$, rather than as an unbiased estimator of θ, yields

$$\alpha^* = \frac{\text{bias}^2(\hat{\theta},\tilde{\theta})}{\text{mse}(\hat{\theta},\tilde{\theta})}.$$

The optimal mse is

$$\text{mse}(\hat{\theta}(\alpha^*),\theta) = \frac{(\tilde{\theta}-\theta)^2 \text{var}(\hat{\theta})}{(\tilde{\theta}-\theta)^2 + \text{var}(\hat{\theta})} = \alpha^* \text{var}(\hat{\theta}). \tag{3.2}$$

The mse is less than $\text{var}(\hat{\theta})$ whenever the simulation has any sampling error.

Notice that binary choice is the special case of linear combination in which α is restricted to being zero or one, which precludes the binary-choice estimator from obtaining the optimal mse. Therefore, if there is any approximation error, the optimal linear combination dominates the optimal binary choice; i.e., $\text{mse}(\hat{\theta}(\alpha^*),\theta) \leqslant \text{mse}(\hat{\theta}^{\text{b}}(\beta^*),\theta)$.

The optimal weight α^* is unknown, but can be estimated. The *natural* estimator of α^* is

$$\hat{\alpha}^* = \frac{(\tilde{\theta}-\hat{\theta})^2}{(\tilde{\theta}-\hat{\theta})^2 + \widehat{\text{var}}(\hat{\theta})} = \frac{\widehat{\text{mse}}(\tilde{\theta},\theta)}{\widehat{\text{mse}}(\tilde{\theta},\theta) + \widehat{\text{mse}}(\hat{\theta},\theta)},$$

which yields the *natural linear-combination* point estimator

$$\hat{\theta}(\hat{\alpha}^*) = \hat{\theta} + \frac{\widehat{\text{var}}(\hat{\theta})(\tilde{\theta}-\hat{\theta})}{(\tilde{\theta}-\hat{\theta})^2 + \widehat{\text{var}}(\hat{\theta})}.$$

Fig. 5 compares the mse of six linear-combination estimators. As in Fig. 1, the axes are in units of $\text{var}(\hat{\theta})$, the 45° line corresponds to the approximation $\tilde{\theta}$, and the horizontal line corresponds to the simulation estimator $\hat{\theta}$. The other two straight lines, corresponding to $\alpha = \frac{1}{4}$ and $\alpha = \frac{1}{2}$, indicate the continuum between the approximation at $\alpha = 0$ and the simulation at $\alpha = 1$. The two curves correspond to the optimal (but unimplementable) linear combination $\hat{\theta}(\alpha^*)$ and the natural linear combination $\hat{\theta}(\hat{\alpha}^*)$, with the optimal curve being the lower. (The four lines are from Eq. (3.1), the lower curve is from Eq. (3.2), and the upper curve is based on a Monte Carlo experiment with negligible sampling error and assuming independent normally distributed samples of size 30.)

The (unimplementable) optimal linear combination $\hat{\theta}(\alpha^*)$ provides a lower mse bound for all linear combinations that use an estimated value of α. The α^* curve is the locus of solutions to the problem of minimizing mse for a given approximation squared error $(\tilde{\theta}-\theta)^2$. This optimal curve is tangential to the constant-α lines, asymptotically approaching the horizontal simulation line $\text{mse}(\hat{\theta},\theta)$ as the approximation error increases. The smaller the value of α, the closer to the origin is its tangent point; the tangent point is at the origin for a perfect approximation and moves toward infinity as the approximation error increases.

Other than the optimal α^* curve, no choice of α dominates the others. Small values of α are good when the approximation error is small.

Now consider the performance of the natural linear combination, $\hat{\theta}(\hat{\alpha}^*)$. In the best case, $\tilde{\theta} = \theta$ and the natural linear-combination estimator reduces mse to about one half that of $\hat{\theta}$. If $\tilde{\theta}$ is within about 1.5 standard errors of θ, then the natural linear combination performs better than $\hat{\theta}$. (Notice that 1.5 standard errors corresponds to 2.25 $\text{var}(\hat{\theta})$.) If the approximation error is greater than 1.5 standard errors, then the mse is larger than that of $\hat{\theta}$, with the maximal 25% increase occurring when $\tilde{\theta}$ is about three standard errors from θ. (Three standard errors lies off the plot to the right at $(\tilde{\theta}-\theta)^2 = 9\,\text{var}(\hat{\theta})$.) Beyond about five standard errors, $\text{mse}(\hat{\theta}(\hat{\alpha}^*,\theta) \approx \text{mse}(\hat{\theta},\theta)$, because the linear combination is giving little weight to $\tilde{\theta}$.

As with Figs. 1 and 2 for the binary-choice estimator, the performance of the linear-combination estimator can be viewed in a second, equally useful, way. Fig. 6 contains the same information as Fig. 5, with mse shown as a function of the simulation mse, both scaled by the approximation mse. As in Fig. 5, every value of α corresponds to a straight line tangent to the optimal curve, but now the slope is α^2 rather than $(1-\alpha)^2$. Simulation alone ($\alpha = 1$) is the 45° line; approximation alone ($\alpha = 0$) is the horizontal line.

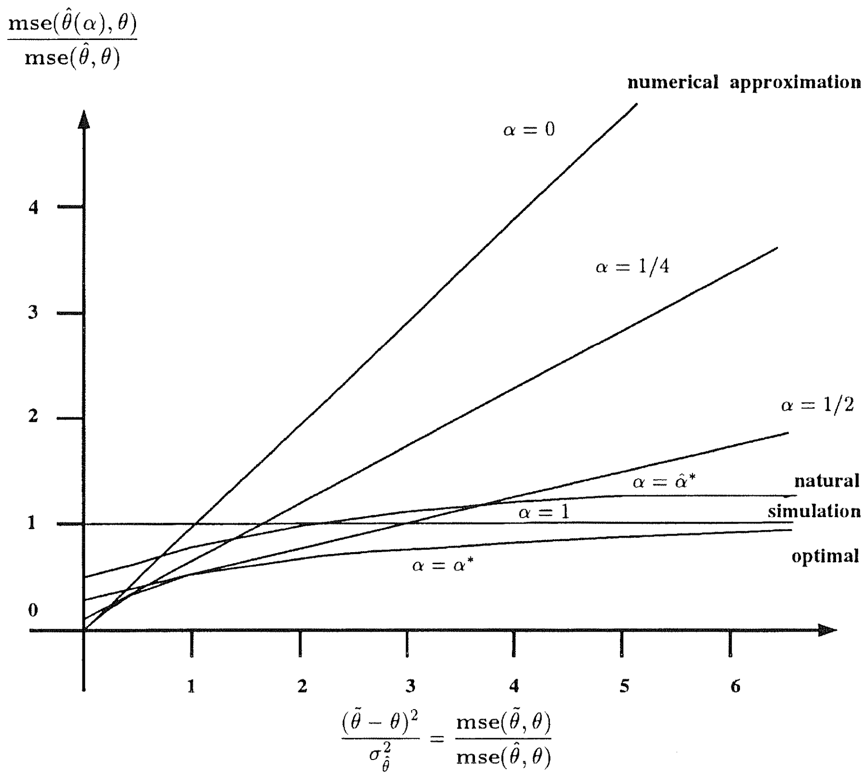

Fig. 5. For various parameter values α, the linear-combination estimator mse as a function of the approximation mse, both scaled by the simulation mse.

The simple average ($\alpha = \frac{1}{2}$) and the theoretical optimal ($\alpha = \alpha^*$) curves appear unchanged. The natural estimator ($\alpha = \hat\alpha^*$) is visually quite different from its curve in Fig. 5. Corresponding to its intercept being one-half in Fig. 5, its slope is asymptotically one-half; that is, when the approximation is much better than the simulation, the natural linear-combination estimator has half the mse of the simulation estimator.

From either Fig. 5 or Fig. 6, performance comparisons can be made. The natural estimator is better than both the simulation and the approximation alone whenever (approximately) $0.47 < \mathrm{mse}(\hat\theta,\theta)/\mathrm{mse}(\tilde\theta,\theta) < 1.4$. The intuitively appealing simple average ($\alpha = \frac{1}{2}$) is better than the simulation, the approximation, and the natural estimator whenever (approximately) $\frac{1}{3} < \mathrm{mse}(\hat\theta,\theta)/\mathrm{mse}(\tilde\theta,\theta) < 3$, but (unlike the natural estimator) performance is arbitrarily bad in the limit as either the simulation or the approximation becomes good, in which case simulation alone or approximation alone is optimal.

4. The Bayesian analysis

Now assume that in addition to the observed values $\tilde\theta$, $\hat\theta$, and $\hat\sigma_{\hat\theta}$, the analyst has available $\sigma_{\tilde\theta-\theta}$, the standard deviation of the Bayesian prior distribution of the approximation error $\tilde\theta - \theta$. More specifically, the approximation error (prior to seeing the simulation estimator $\hat\theta$) is now modeled as a normal random variable with mean zero and variance $\sigma^2_{\tilde\theta-\theta}$.

Such prior information about the quality of the approximation is often unavailable, but analyst experience or approximation bounds sometimes suggest a reasonable value of $\sigma_{\tilde\theta-\theta}$. The random approximation error is equivalent to expressing uncertainty about θ: having $\tilde\theta$ in hand, the analyst views the

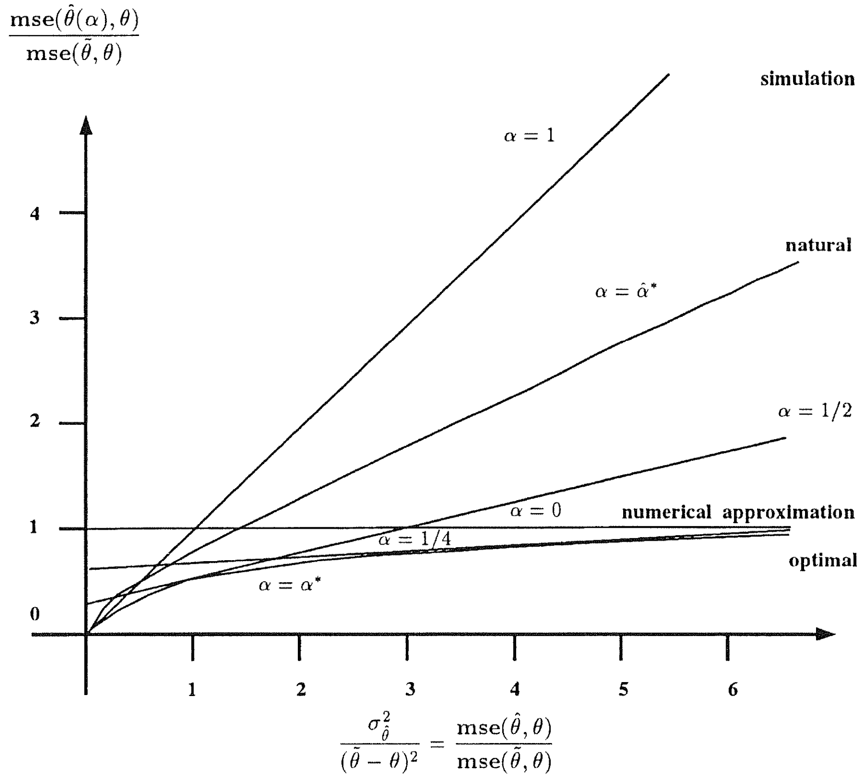

Fig. 6. For various parameter values α, the linear-combination estimator mse as a function of the simulation mse, both scaled by the approximation mse.

unknown θ as a random variable with mean $\tilde{\theta}$ and standard deviation $\sigma_\theta = \sigma_{\tilde{\theta}-\theta}$.

If the simulation estimator is normally distributed, the standard Bayesian analysis (e.g., Berger, 1985, p. 128) yields a normal posterior distribution whose mean (and therefore a reasonable point estimator) is the linear-combination $\hat{\theta}(\alpha') = \alpha'\hat{\theta} + (1-\alpha')\tilde{\theta}$ with simulation weight

$$\alpha' = \frac{\sigma^2_{\tilde{\theta}-\theta}}{\sigma^2_{\tilde{\theta}-\theta} + \sigma^2_{\hat{\theta}}}.$$

The natural estimator of α' is

$$\hat{\alpha}' = \frac{\sigma^2_{\tilde{\theta}-\theta}}{\sigma^2_{\tilde{\theta}-\theta} + \hat{\sigma}^2_{\hat{\theta}}}.$$

The mse of this Bayesian linear combination, $\hat{\theta}(\hat{\alpha}')$, is identical to the mse of the constant-α linear combinations shown in Fig. 5. Because the units are var$(\hat{\theta})$, the transformation from prior variance to Fig. 5 is $\alpha = \sigma^2_{\tilde{\theta}-\theta}/(\sigma^2_{\tilde{\theta}-\theta} + 1)$ and the transformation from α in Fig. 5 is $\sigma^2_{\tilde{\theta}-\theta} = \alpha/(1-\alpha)$. For example, the $\alpha = \frac{1}{4}$ line in Fig. 5 is also the $\sigma_{\tilde{\theta}-\theta}/\sigma_{\hat{\theta}} = \frac{1}{3}$ line; that is, the analyst has specified a prior standard deviation that is one-third of the simulation's standard error $\sigma_{\hat{\theta}}$. As the prior standard deviation decreases, the Bayesian linear-combination line moves closer to the 45° approximation line.

5. Summary and discussion

All three alternatives – binary choice, linear combination, and Bayesian analysis – for combining approximation and simulation estimators shift mse.

The simulation mse is reduced when the approximation is good, and increased when the approximation is poor. Equivalently, the approximation mse is reduced when the simulation has small standard error, and increased when the simulation has large standard error.

Neither of the first two alternatives, which have no specification of approximator quality, guarantees an improved estimator. In the best case – when the approximation has no error – mse is decreased by the natural binary-choice estimator by only about 20% and by the natural linear-combination estimator by about 50%. In the worst case – when the approximation has an error of several simulation standard errors – the mse is increased by about 25%. When the approximation error is asymptotically infinite, both natural methods ignore the approximation and therefore have the simulation mse.

The third alternative, which requires specification of a Bayesian prior on the approximation error, can reduce mse to zero in the best case – when the approximation error is zero and the prior variance is zero – but can increase mse by arbitrarily large factors when the prior variance is misleading. This alternative, unlike the first two, fails to converge to the simulation mse as the approximation error increases.

The binary-choice alternative has little to recommend it. Despite its intuitive appeal and its wide informal use, the natural binary-choice mse is only slightly better than the simulation mse, and then only when the squared approximation error is quite small: less than about one-half of the simulation variance.

The linear-combination alternative is more appealing than binary choice, which is a special linear combination. The simulation mse is reduced whenever the squared approximation error is less than twice the simulation variance; the approximation mse is reduced whenever the squared approximation error is greater than about one-half of the simulation variance. The natural linear-combination estimator sometimes (squared approximator error ranging from one-half to two simulation variances) does better than the optimal binary-choice estimator. Despite a slightly larger mse when the approximation error is large, as a practical matter the natural linear-combination dominates the natural binary-choice estimator.

Whether the potential mse reduction is worthwhile depends on the application. The maximal 50% mse reduction corresponds to a 50% reduction in computer run time. Whatever the reduction, it is easy to obtain and essentially free, costing only the approximation computation. For queueing systems, approximations are often most accurate in heavy traffic, which is where simulation is least efficient. Thus the requirement of the approximation being within two standard errors is frequently met. Therefore, the best mse reduction is likely to occur in applications where it is most needed.

The linear-combination approximation-assisted point estimator of Section 3 is a special case of using a biased external control variate (Schmeiser and Taaffe, 1994). In the notation and terminology of that paper, the control-variate estimator is $\hat{\theta}^c(\alpha) = \hat{\theta} - (1-\alpha)(\hat{\theta}^a - \tilde{\theta}^a)$, with $\hat{\theta}$ and $\hat{\theta}^a$ being positively correlated simulation point estimators of the principal model and an approximation model, respectively, and $\tilde{\theta}^a$ being a numerical approximation to $\theta^a = \mathrm{E}(\hat{\theta}^a)$, the true performance of the approximation model. If we assume that the approximation model is identical to the principal model, and therefore that $\theta^a = \theta$, then the approximation-model simulation is unnecessary by setting $\hat{\theta}^a = \hat{\theta}$ and the approximation-model numerical approximation becomes $\tilde{\theta}^a = \tilde{\theta}$. The control-variate estimator then simplifies to the linear-combination approximation-assisted point estimator $\hat{\theta}^c(\alpha) = \alpha\hat{\theta} + (1-\alpha)\tilde{\theta}$.

References

Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer, New York.

Bickel, P. and K.A. Doksum (1977), *Mathematical Statistics*, Holden-Day, San Francisco, CA.

Schmeiser, B. and M. Taaffe (1994), Time-dependent queueing network approximations as simulation external control variates, *Oper. Res. Lett.* **16**, 1–9.

Whitt, W. (1983), The queueing network analyzer, *Bell Systems Tech. J.* **62**, 2779–2815.